

CITAD 8 PRE-PROCEEDINGS

Eighth International Conference on the Anthropological Theory of the Didactic



19-23 January 2026

Barcelona

Spain

An Artificial Intelligence (AI) approach for assisting teachers with the implementation of study and research paths

Veronica Parra^{1,2,3}, Ana Corica^{1,2,3}, Silvia Schiaffino^{1,3}, Daniela Godoy^{1,3}, Berta Barquero⁴, Marianna Bosch⁴, Marc Comas⁵

¹ Consejo Nacional de Investigaciones Científicas y Técnicas (CONICET), Argentina;

² Núcleo de Investigación en Educación Matemática (NIEM-CICPBA), UNCPBA, Tandil, Argentina, [vparra,acorica}@niem.exa.unicen.edu.ar](mailto:{vparra,acorica}@niem.exa.unicen.edu.ar)

³ Instituto Superior de Ingeniería de Software (ISISTAN - CONICET-UNCPBA), Tandil, Argentina [silvia.schiaffino,daniela.godoy}@isistan.unicen.edu.ar](mailto:{silvia.schiaffino,daniela.godoy}@isistan.unicen.edu.ar)

⁴ Dep. d'Educació Lingüística, Científica i Matemàtica, Universitat de Barcelona, Spain [marianna.bosch,bbarquero}@ub.edu](mailto:{marianna.bosch,bbarquero}@ub.edu)

⁵ Dep. d'Informàtica, Matemàtica Aplicada i Estadística, Universitat de Girona, Spain, marc.comas@udg.edu

This paper proposes the application of Artificial Intelligence (AI) techniques to assist teachers during the implementation of study and research paths (SRPs). The use of AI allows the analysis of derived questions in a class to create a report that summarises the class activity. This report organises the multiple and varied questions that might have emerged into automatically discovered topics using an unsupervised machine learning approach. Natural language processing (NLP) is applied to students' questions in order to semantically group them into meaningful topics. The proposed AI method is seemingly integrated into a Web platform designed to accompany and digitalise the entire SRP implementation process. The ultimate goal of the proposed AI support is to alleviate the workload of teachers from one class to the next, allowing them to focus on the design of the important aspects of the following class.

Keywords: Artificial intelligence, web platform, study and research paths, inquiry management.

Introduction

Study and research paths (SRPs) (Chevallard, 2019) are instructional proposals elaborated within the Anthropological Theory of the Didactic (ATD) to implement inquiry processes as normalised activities in the classroom. The starting point of an SRP is a generating question Q formulated in a strong sense, that is, with a non-immediate answer, generating the need to raise sub-questions, called derived questions. Thus, knowledge (mathematical and non-mathematical) will be studied in order to construct answers to such questions, shifting toward a more functional study of mathematics than the one proposed in official programs. Questions such as: *Why do bugs go toward the light?* (Chevallard, 2009), *How much benefit do producers and consumers receive from the existence of a competitive market?* (Benitez & Parra, 2024), or *Has the creation of "low emissions zones" reduced pollution?* (Fernández-Ruano et al., 2024, Verbisck et al., 2024) are examples of questions generating study and research paths (SRPs).

Constructing answers to a generating question is neither a linear nor a predetermined process. Within a classroom, students (organised in small teams) can pursue their own paths. In consequence, the derived questions will not be the same for each group, nor will the resources used to search for answers. This type of teaching requires radical changes in the roles of teachers and students (*topogenesis*), in the management of the inquiry milieu (*mesogenesis*), and in the teaching time (*chronogenesis*) required to develop an SRP, since it is opposite to the scheme of the teacher

explaining and the students listening, copying, and repeating exercises. SRPs involve simulating, within the classroom, the work of a research team when addressing open problems. In these cases, researchers do not know the solution in advance; they must find it collaboratively, researching and studying together. In the classroom, the teacher must guide the inquiry based on the progress of each student group, without necessarily being able to determine in advance where the process will lead. Even if the teacher conducts an a priori analysis of the possible paths and the knowledge these paths can lead to, there is no certainty that what has been analysed beforehand will actually happen. Furthermore, the teacher must ensure that students engage with the mathematical knowledge that helps satisfy the needs of the inquiry.

Managing SRPs, therefore, implies that the mathematics teacher must assume a different role in comparison with traditional teaching. This is a teacher who supervises the students' activities, who do not necessarily follow the same path and complete the same studies in a class. The path followed by a group depends on the type of question generated by each student or team of students, which ultimately enables them to provide answers to the derivative question. This type of teaching requires students to formulate questions and provide answers using a variety of resources. At the same time, because all the class addresses the same generating question, the inquiry process needs to regularly pool and discuss the results obtained by each team of students within the class group. As the inquiry guide, the teacher has a critical responsibility in managing these discussions and helping make decisions about the questions to discard and the paths to follow.

SRPs implementation requires the teacher to allocate significant time between classes, as each class is designed based on the students' achievements of the previous one. In other words, not all classes can be fully planned from the beginning, as it is a sequence where the next class is planned based on the progress made in the previous one. This requires the teacher to analyze, synthesize, and make decisions from one lesson to the next about how to guide and advance the inquiry proposed by each group. This, however, demands time from teachers, which is scarce due to the demands of teaching multiple courses, possibly in several institutions.

To offer a solution to this problem, this work proposes the application of Artificial Intelligence (AI) techniques, specifically machine learning (ML) and natural language processing (NLP), to provide intelligent assistance to teachers during the implementation of a SRP. The final goal of using AI at this stage is to automatically generate a report showing how the derived questions discussed in class can be grouped into topics that are automatically discovered through the application of unsupervised machine learning, thus summarizing the work made in class. This provides the teacher with a summary of each group's work after every class session, saving time and enabling informed decisions for steering the inquiry in the next session.

AI-based automatic processing of questions after finishing a class seeks to emulate the process typically done manually by teachers: reading the students' diaries and organising the questions that arose in class to identify the commonalities and specificities, before deciding how to continue in the subsequent class. Having an initial classification of the derived questions powered by AI allows a quick visualization of the issues covered, and the associated questions considerably optimises this stage of SRP implementation for teachers. Thus, the time previously spent on manual analysis can shift towards the management of the next class, deciding about whether to intervene so that different teams can reformulate questions, generate new ones, delve deeper into the answer to a question or undertake new searches for resources to provide answers.

The paper is organised as follows. First, the type of report generated by AI is illustrated. Then, the methods used in its generation are detailed, including the processing of natural language questions and the machine learning approach. Finally, the empirical evaluation of the approach taking the

work of teachers as external reference is summarised. The paper closes with some concluding remarks and future lines of research.

AI-support for class report generation

As previously explained, the goal of the incorporation of AI in the process of implementation of a SRP is to alleviate the teacher’s workload, including the additional time required after each session to sort, analyse, and evaluate the potential and limitations of the questions agreed upon by each small group, and to make decisions that allow timely intervention so that the study can progress.

Elements of a class report generated with AI

At the end of an SRP session, the teacher must analyse multiple questions that arose from different teams and that possibly delve into different aspects of the generating question under study. We propose an AI-based method that produces a report identifying these different aspects or topics raised by the students during the session, allowing a quick overview of the questions covered in class. Figure 1 illustrates an example of a report generated by this method.

The report summarises how the questions can be grouped into a number of topics automatically discovered from the data (derived questions that emerge during the class). The figure shows a set of derived questions grouped in multiple topics, each associated with a frequency (votes), of questions from a SRP related to the delimitation of low emissions zones (ZBE in Catalan) in Catalonia cities (Fernández-Ruano et al., 2024; Verbisck et al., 2024). A total of 45 questions formulated by the six teams of students were organised by the algorithm into three large groups. Some questions appeared in more than one group, and as it can be seen in the figure, and some even arose in all groups. Thus, the report quickly identifies repeated questions and groups them in these three broad topics.

Preguntas por temas

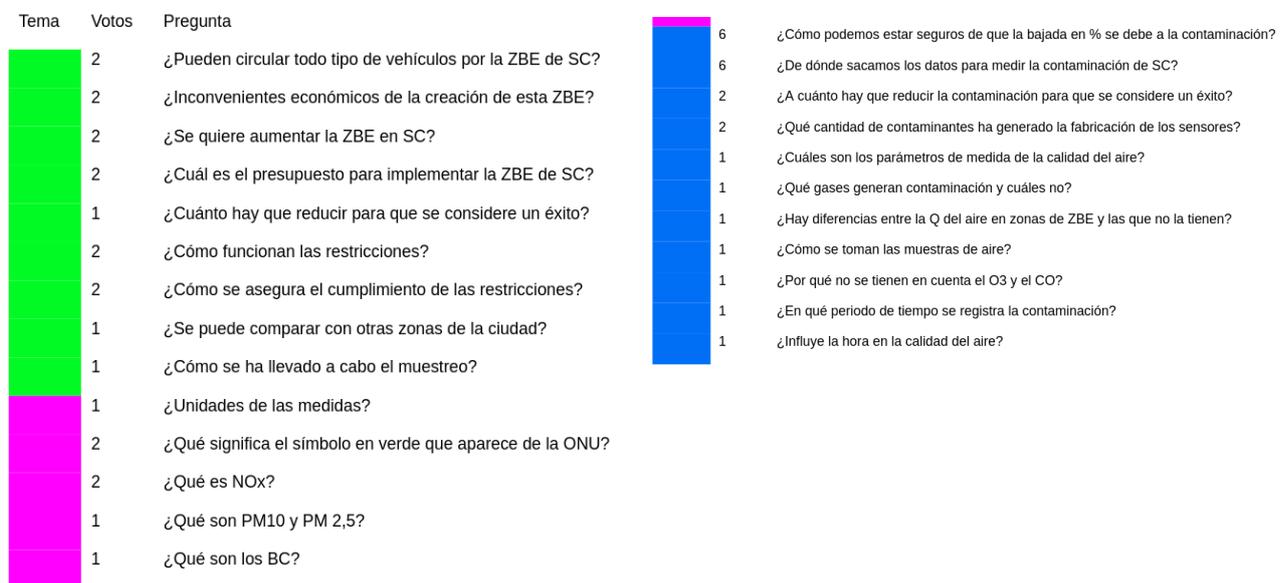


Figure 1: Discovered topics and summary of questions

For each of the identified topics (three in this example), the report also provides a word cloud (Figure 2), to visualise the prevalence of words in each of the topics.

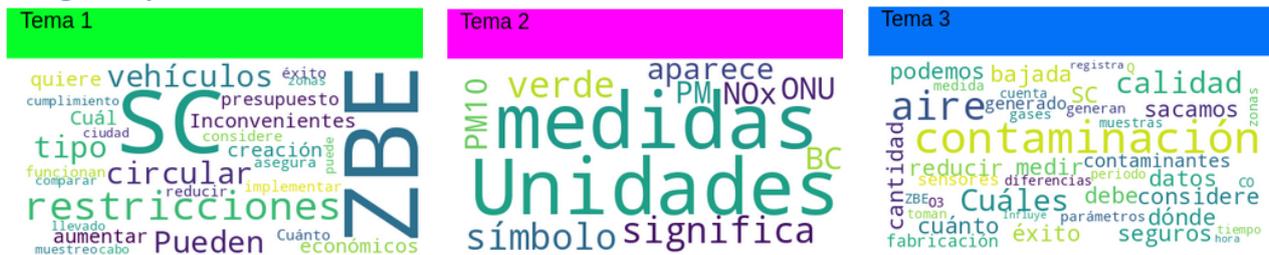


Figure 2: Word clouds associated to each of the three topics in the SRP

A further summary also shows the questions within the topic, indicating the student teams that asked them (Figure 3). This figure presents Topic 1 as an example. First, we can see the word cloud representing the topic. Next, a matrix is shown where the first column indicates the number of times the question was detected. The second column lists the group of students who asked the question (in this case, the students' anonymity has been maintained). Finally, the third column lists the question itself. This provides a description of the questions asked, which group asked them, and how many times the question arose.

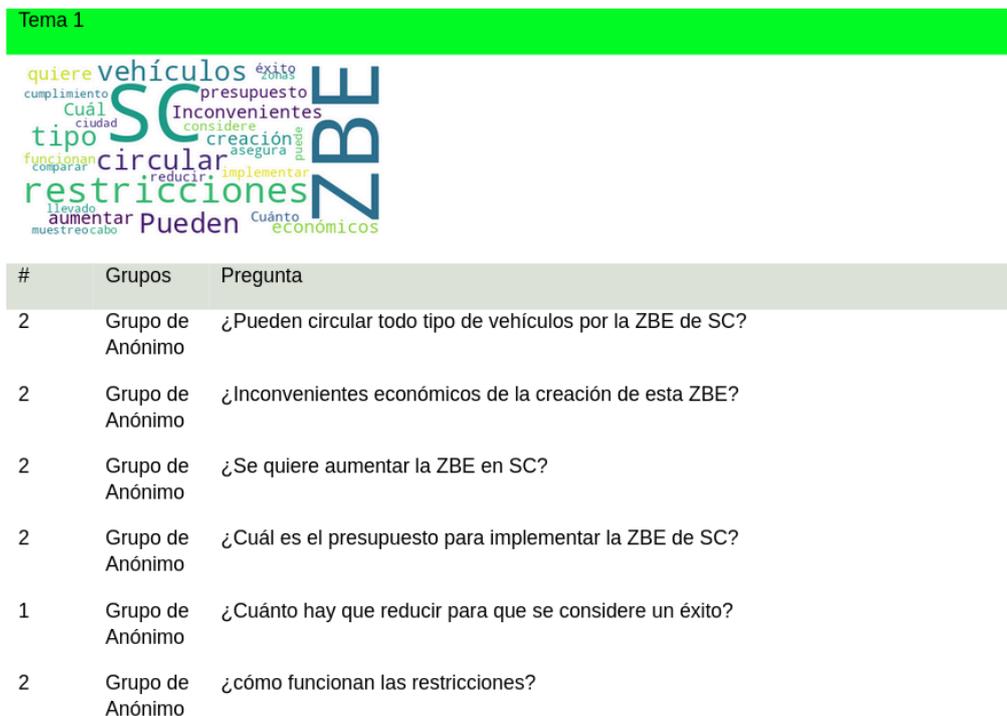


Figure 3: Description of a topic discovered using AI

Unsupervised machine learning for topic discovery

From the perspective of an AI algorithm, creating a report like the one shown in the previous section involves taking a set of questions expressed in natural language as input and discovering the themes or topics they belong to. Figure 4 provides a general overview of the entire process towards accomplishing this goal, which steps are described in detail in this section.

Since the set of questions constitutes unlabelled data, lacking a prior classification or categorisation, the learning task falls within the umbrella of unsupervised machine learning, specifically clustering

methods. Clustering algorithms are a type of ML algorithms aiming at organising or categorising different objects, data points, or observations into groups or clusters based on their similarity. Each group or cluster is composed of objects that exhibit greater similarity to one another than to those in other clusters. In the absence of a priori knowledge about the classes or categories of the questions under analysis, unsupervised clustering methods seem to be ideal for organising questions into meaningful groups or clusters.

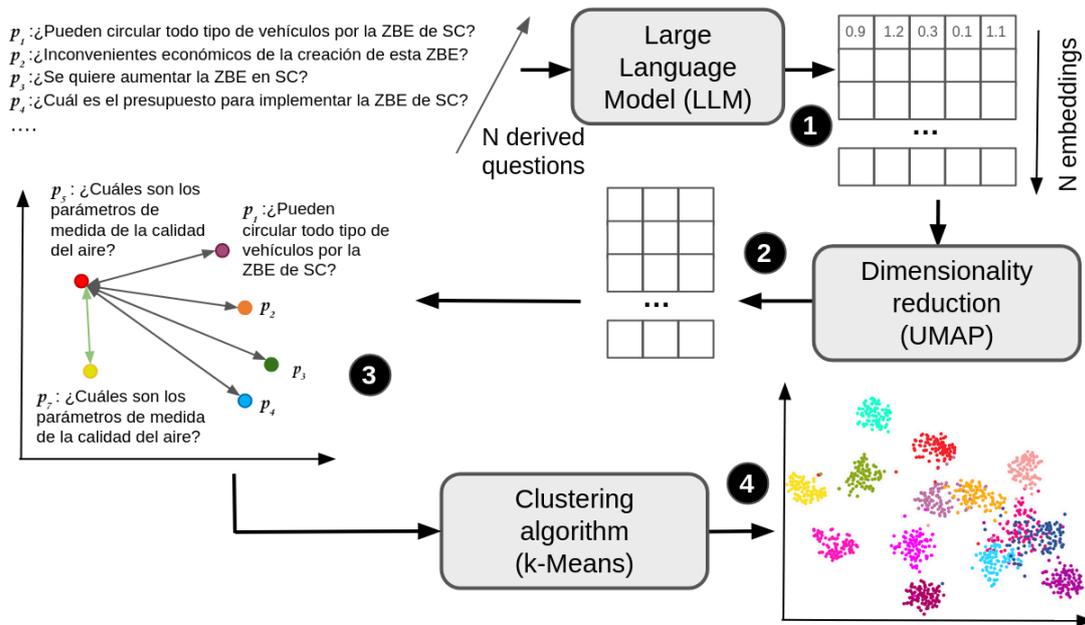


Figure 4: Overview of the topic discovering based on unsupervised ML

Clustering algorithms are guided by the notion of similarity (or distance) between objects (questions in this case). In the case of natural language texts, this similarity should capture the semantics of words rather than its syntax. Thus, to be processable for a machine learning algorithm, natural language questions have to be transformed into a computational text representation that captures their meaning. Current state-of-the-art NLP techniques are based on projecting a piece of text into a high-dimensional latent space of *embedding* vectors, or embeddings for short. In this space, the proximity between vectors reflects semantic similarity.

Embedding vectors are obtained using pre-trained models that capture word relationships and language nuances (Fig. 4, step 1). In this high-dimensional space, words with similar meanings are located close to each other, while words with different meanings are placed farther apart. Questions to be clustered in this work consist of a set of words that need to be interpreted not only by their own semantic meaning but also within the context of the sentence. For this purpose, we use *Sentence transformers*¹ (a.k.a. Sentence-BERT or SBERT) (Reimers and Gurevych, 2019), a type of deep learning model specifically designed to capture the semantic meaning of sentences, going beyond the capabilities of traditional word embeddings. SBERT encodes entire sentences or text fragments into fixed-size vectors, preserving the contextual meaning. These models are used to bring semantically similar sentences closer together in the embedding space, while separating dissimilar ones (Fig. 4, step 3). The training process of models uses natural language inference data, typically involving pairs of sentences labelled as similar or dissimilar. This approach has

¹ <https://sbert.net/>

significantly enhanced the performance of NLP applications in the last few years, enabling more accurate and meaningful text analysis. In order to make this method widely applicable for implementing SRP in different languages, a multilingual model such as *MiniLM-L12-v2*² is applied.

General-purpose pre-trained models such as *MiniLM-L12-v2* produce high-dimensional embedding vectors. UMAP (Uniform Manifold Approximation and Projection) (McInnes & Healy, 2018) is used as an intermediate step (Fig 4, step 2) for reducing the space dimensionality while preserving the underlying structure of the data. This reduction often reveals interpretative insights hidden in noisy and sparse data.

Clustering of questions is carried out in the resulting lower-dimensional space (Figure 4, step 4). We used a classical partitioning clustering algorithm as *k*-Means, whose goal is to partition *n* observations (questions in this case) into *k* clusters. This algorithm relies on the notion of centroids, where each centroid represents the centre of a cluster, and it is essentially the mean or average of all data points assigned to that cluster. Thus, the algorithm groups similar data points together, minimizing the distance between each point to its assigned cluster centroid.

One of the known limitations of *k*-Means is that it requires the number of clusters (*k*) to be specified beforehand. To circumvent this limitation, heuristics methods are frequently used, including the elbow method, silhouette or gap statistics. In this work, we used the Calinski-Harabasz (CH) index (Caliński & Harabasz's, 1974), also known as the Variance Ratio Criterion (VRC), as a method to evaluate the goodness of a partition given by a number of clusters. This index is calculated as a ratio of the sum of inter-cluster dispersion and the sum of intra-cluster dispersion for all clusters (where dispersion is the sum of squared distances). A high CH means better clustering, since observations in each cluster are closer together (more cohesive), while clusters themselves are further away from each other (well separated). The procedure involves running *k*-Means for different values of *k* (from 2 to a predetermined maximum), calculating the CH index for each case. Finally, the *k* value that maximises the index is selected, as this indicates a well-defined and separated cluster structure. This approach allows the number of clusters to be tailored to the inherent structure of the data, without requiring prior knowledge of the number of underlying classes.

Figure 5 shows a visualisation in a two-dimensional space (reduced with UMAP) of a set of question groups obtained after a class once the complete process has been run: representing the text with embeddings, reducing dimensionality, and executing the clustering algorithm. As can be seen in the figure, the groups reflect emerging themes or topics from the class discussion.

² <https://huggingface.co/sentence-transformers/all-MiniLM-L12-v2>

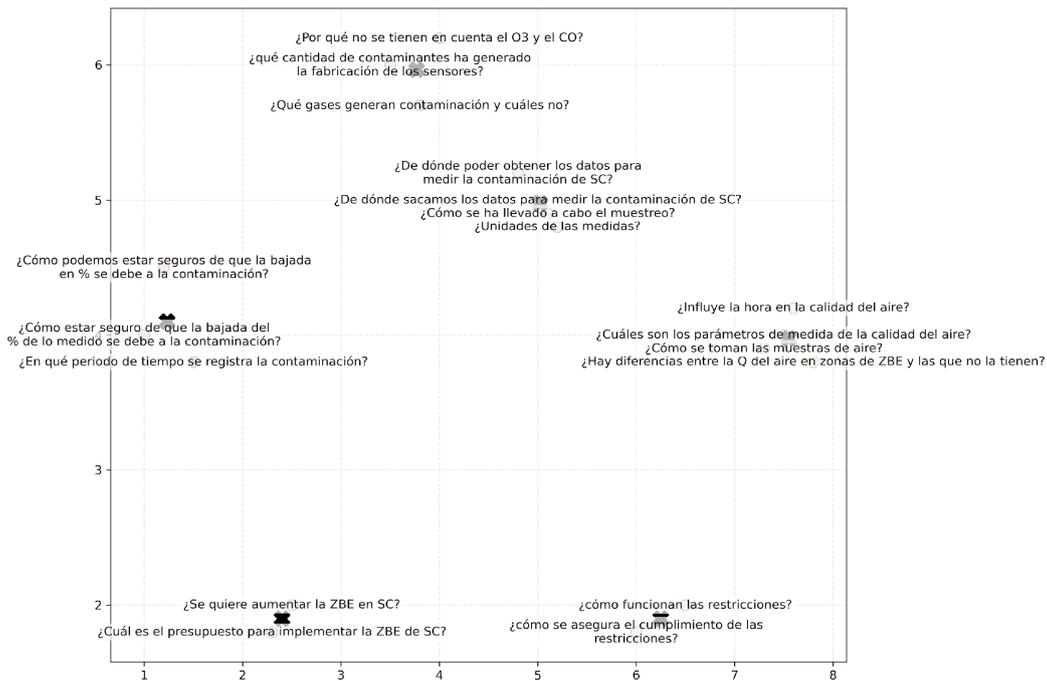


Figure 5: Visualization of questions clustering (Paladino Bravo, 2025)

Empirical validation of clustering results

Evaluation of the validity, quality or accuracy of clustering solutions is a difficult issue since clustering is an exploratory task within ML that aims at knowing more about the data. In consequence, there may be multiple different but equally good and valid clustering solutions. In this regard, clustering can be seen as a somewhat subjective procedure in which groups can be formed based on diverse criteria. It is worth noticing that this also applies to the task of grouping questions when it is done by teachers. For the exactly same set of questions, different teachers will use their own grouping criteria, which might result in different groupings.

Measures of cluster validity can be divided into two types: internal and external (Jain and Dubes, 1988). Internal validity criteria compare different sets of clusters without reference to external knowledge as a function of the given data and/or similarities. External validity criteria measure how well the clustering results match some prior knowledge about the data. The most common form of external information is a set of classes and class labels for instances usually obtained by manual classification.

In the context of this work the external validation should be done by teachers, which are the targeted end-users of the final AI report. To conduct this evaluation, we asked four teachers to manually classify a set of 94 real questions posed during a REI regarding low-emission zones in Girona. Each question was then labelled with multiple reference classifications (known in ML as ground truths) proposed by the teachers. Question labels assigned by teachers express different interpretations of how the questions should be grouped according to the criteria of the four teachers involved, denoted JB, KM1, NR1, and MB. The answers of these four teachers were chosen for the experiments as they were the ones that assigned categories to the complete list of 94 questions.

It should be noted that, in some cases, teachers proposed hierarchical classifications with varying levels of granularity. For example, Teacher KM classified using a general category (KM1) and a more specific subcategory (KM2), while Teacher NR used up to three levels (NR1, NR2, and NR3). To maintain the feasibility of the analysis and avoid an explosion in the number of expected

clusters, the evaluation was conducted considering the most general level (KM1 and NR1). Figure 6 shows the distributions of the classes or categories that each teacher assigned to the questions. As can be seen in the graphs, the categories defined by each of them differ significantly. For example, MB organized the questions into 7 groups, while KM1 organized them into 12. In addition to the discrepancy in the number of classes (which are the expected clusters), there is a noticeable imbalance in the number of questions in each class, which represents an additional challenge for the clustering algorithm.

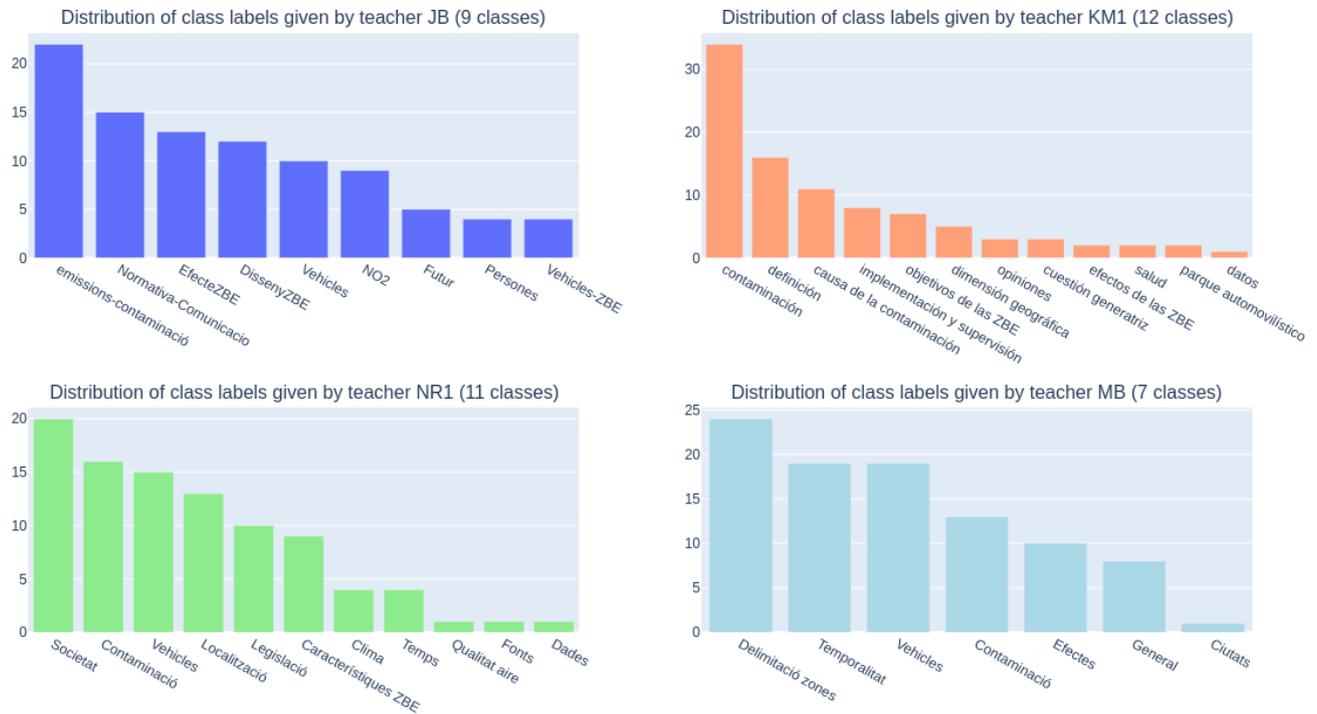


Figure 6: Labels distribution in the classifications made by a group of teachers for a set of questions.

External validation clustering metrics are then used to evaluate the consensus between the clustering solution provided by the algorithm and that given by a teacher. We used Adjusted Mutual Information (AMI), Normalized Mutual Information (NMI), Purity, and V-measure (Jain & Dubes, 1988) to compare clustering results with human-made reference labels (ground truth), thus assessing the resulting grouping validity. Table 1 reports the results obtained of number of clusters (#clusters) and validity measures when comparing the clustering algorithm and the classification made by the four teachers (JB, KM1, MB, NR1). The best results are indicated in bold.

Teacher	# clusters	AMI	NMI	Purity	V-Measure
JB	12.46 ± 0.90	0.38 ± 0.01	0.52 ± 0.01	0.61 ± 0.01	0.52 ± 0.01
KM1	12.46 ± 0.90	0.12 ± 0.01	0.35 ± 0.01	0.48 ± 0.01	0.35 ± 0.01
MB	12.46 ± 0.90	0.30 ± 0.01	0.43 ± 0.01	0.61 ± 0.01	0.43 ± 0.01
NR1	12.46 ± 0.90	0.24 ± 0.01	0.42 ± 0.01	0.49 ± 0.02	0.42 ± 0.01

Table 1: Comparison of clustering results and teachers' manual classifications.

In terms of the agreement of the partitions produced by the clustering algorithm with those made by teachers, we can observe that the algorithm produced in a higher number of clusters, but there is almost no variation for different models and different executions (denoted by the standard deviation). Teachers divided the examples in 9, 12, 7 and 11 groups, while the algorithm discovered

about 12 clusters, with a minimum of 11 and a maximum of 16. Beyond the number of clusters, it is important to assess the goodness of the semantic groupings. In this regard, the scores show a good performance of the clustering algorithm, again considering the difficulty and highly subjective nature of the clustering task at hand. As can be seen in the table, the algorithm produced clusters closest to the classification made by teacher JB, and secondarily by teacher MB. The most distant classification was that of teacher KM1, which appears to follow a non-thematic criteria.

Concluding remarks

In this work, we proposed a novel AI-based approach for assisting in the analysis of derived questions formulated in class sessions during the implementation of an SRP. The goal is to bring intelligent assistance to teachers implementing SRPs by providing an automatic report with an initial organisation of the questions, which can be numerous and varied, thus reducing the time required to read and organise them manually. This constitutes a resource that can save the teacher's time, although it does not exempt her from carrying out an analysis and evaluation of the report obtained in order to make decisions for the next session with her students. Based on advanced NLP techniques, questions expressed in natural language are grouped into topics or themes providing a quick visualisation and favouring interpretation, allowing the teacher to work on them further, saving a significant amount of time that can be better devoted to designing the next lesson. It is not about replacing the teacher's work, but about having an assistant that can provide an initial analysis of what happened in the class.

The proposed AI method described in this work is accessible through a Web platform (<https://reis.apps.isistan.unicen.edu.ar/>) exclusively designed to accompany and digitalise the entire SRP implementation process (Corica et al., 2025). The platform allows students to easily load the questions they are working on in the teams, the answers they found, the sources used and the way each team worked, among other things, using any device (phone, computer, or tablet). For that, the teacher must first create a new SRP in the platform and the corresponding sessions. Upon completion of a class, the teacher receives a report as the one described above, both in the platform and by email. The platform closes the circle of implementation of an SRP so that all the process is registered and can be studied as a whole for obtaining insights for future implementations.

Although the use of AI described in this paper is oriented to assist the teacher in a given SRP session summary, the used techniques, together with the platform that stores all the information generated by the inquiry, enable extending the intelligent assistance to other points of the SRP process. In future research, we will explore the use of these technologies to track and map the evolution of an SRP in terms of the set of derivative questions and the branches they open.

Acknowledgment

This work has been partially funded by projects LABINQUIRY (PDC2022-133812-C21, PDC2022-133812-C22, MICIU/AEI/EU/PRTR); EDU-Paradigms (PID2021-126717NB-C31, MCIN/AEI and FEDER A way to make Europe), and PIP CONICET “*La formación de profesores en matemática: análisis didáctico y cognitivo de recursos en la enseñanza de la matemática*” (11220220100429CO). RESOL-2024-436-APN-DIR#CONICET.

References

Benitez, V., & Parra, V. (2024). Teaching mathematics for economists: partial results of the implementation of a study and research path. *Teaching Mathematics and Its Applications: International Journal of the IMA*, 43, 382-402. <https://doi.org/10.1093/teamat/hrae024>.

- Caliński, T., & Harabasz, J. (1974). A dendrite method for cluster analysis. *Communications in Statistics*, 3(1), 1–27. <https://doi.org/10.1080/03610927408827101>
- Chevallard, Y. (2009). Remarques sur la notion d'infrastructure didactique et sur le rôle des PER. [Remarks about the notion of didactic infrastructure and the role of SRP.] http://yves.chevallard.free.fr/spip/spip/IMG/pdf/Infrastructure_didactique_PER.pdf
- Chevallard, Y. (2019). Introducing the anthropological theory of the didactic: An attempt at a principled approach. *Hiroshima journal of mathematics education*, 12, 71-114.
- Corica, A., Parra, V., Paladino Bravo, I., Schiaffino, S., Zunino, A., & Godoy, D. (2025). AI-based Web platform for supporting teaching by study and research paths. *CITAD 8: 8th International Conference on the Anthropological Theory of the Didactic*.
- Fernández-Ruano, L., Serrano, V., Báguena, J., Cuadros, J., & Sáez, M. (2024). A Practical Approach to an SRP in University Statistics. In I. Florensa, N. Ruiz-Munzón, K. Markulin, B. Barquero, M. Bosch & Y. Chevallard (eds.), *Extended Abstracts 2022: Proceedings of the 7th International Conference on the Anthropological Theory of the Didactic (CITAD7)*. Trends in Mathematics (pp. 267-281). Springer. https://doi.org/10.1007/978-3-031-55939-6_21
- Jain, A. K. & Dubes, R. C. (1988). *Algorithms for Clustering Data*. Prentice Hall.
- McInnes, L., & Healy, J. (2018). UMAP: Uniform Manifold Approximation and Projection. *Journal of Open Source Software*, 3(29), 861. <https://doi.org/10.21105/joss.00861>
- Paladino Bravo, I. (2025). *Plataforma web para asistir en la enseñanza por REIs (Recorrido de Estudio e Investigación) en el aula*. [Trabajo Final de Ingeniería en Sistemas, Universidad Nacional del Centro de la Provincia de Buenos Aires]. Argentina.
- Reimers, N. & Gurevych, I. (2019). Sentence-BERT: Sentence Embeddings using Siamese BERT-Networks. *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing (EMNLP)* (pp.3973-3983). Association for Computational Linguistics.
- Verbisck, J., Fernández-Ruano, L. & Lombard, N. (2024). A pilot study and research path in statistics at the university level. En A. González-Martín, G. Gueudet (eds.), *Proceedings of the Fifth Conference of the International Network for Didactic Research in University Mathematics* (pp. 513–522). CRM and INDRUM, 2024. <https://indrum2024.sciencesconf.org/>